DOCUMENT RESUME

ED 352 420 TM 019 356

AUTHOR Gallagher, Ann M.
TITLE Sex Differences in Problem-Solving Strategies Used by High-Scoring Examinees on the SAT-M.
INSTITUTION College Entrance Examination Board, New York, N.Y.
REPORT NO CBR-92-2; ETS-RR-92-33
PUB DATE 92
NOTE 41p.
AVAILABLE FROM College Board Publications, P.O. Box 886, New York, NY 10101 ($12).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Algorithms; Classification; College Entrance Examinations; High Achievement; High Schools; High School Seniors; *High School Students; Interviews; *Mathematics Skills; Mathematics Tests; Predictive Measurement; *Problem Solving; Protocol Analysis; Questionnaires; Scores; *Sex Differences; *Student Attitudes; Student Characteristics
IDENTIFIERS High School Juniors; *Scholastic Aptitude Test; *Strategy Choice

ABSTRACT

An item classification scheme developed by A. M. Gallagher (1990) was refined, resulting in a more accurate prediction of sex differences in the mathematical sections of the Scholastic Aptitude Test (SAT). Differential Item Functioning (DIF) procedures for examinees scoring over 650 indicated that the majority of items favoring males required the use of mathematical insight, while all the items flagged as favoring females required standard algorithmic solutions. Structured interviews were conducted with 25 male and 22 female high school juniors and seniors from the examinee group to study differences in strategy use. There was a substantial overlap in strategies used by males and females, but analysis across all items indicated that females were more likely than males to use algorithmic strategies, and males were more likely than females to use insightful strategies. Questionnaire data obtained from the students interviewed indicated a positive relationship for both males and females between SAT mathematical performance and positive attitudes toward mathematics. The use of algorithmic strategies was correlated with negative attitudes toward mathematics. Implications of these and other findings are discussed. Eleven tables present study findings. Four appendixes contain the following: (1) prototypical items for classification categories; (2) problems used for think-aloud protocols; (3) attitude and background questionnaire; and (4) background variables and correlations between SAT mathematics or strategy and questionnaire items. (Author/SLD)

College Board Report
No. 92-2

Sex Differences in Problem-Solving Strategies Used by High-Scoring Examinees on the SAT®-M

Ann M. Gallagher

Sex Differences in Problem-Solving Strategies Used by High-Scoring Examinees on the SAT-M

Ann M. Gallagher

College Board Report No. 92-2
ETS RR No. 92-33

College Entrance Examination Board, New York, 1992

Ann M. Gallagher is Associate Research Scientist in the Division of Cognitive and Instructional Science at ETS.

Acknowledgments

The author wishes to express her appreciation to Carol Jackson and Jeff Haberstroh of the ETS College Board test development staff for their assistance in revising the item classification scheme and in classifying items.

4

CONTENTS

Tables

ABSTRACT

An item classification scheme developed by Gallagher (1990) was refined, resulting in a more accurate prediction of sex differences in performance on the mathematical sections of the College Board's Scholastic Aptitude Test (SAT). When Differential Item Functioning (DIF) procedures were performed on item data for examinees scoring above 650, the majority of items that were flagged as favoring males required the use of mathematical insight, whereas all the items flagged as favoring females required standard algorithmic solutions.

Structured interviews were conducted with students (25 males and 22 females) in this score range to determine the nature of differences in strategy use. A classification scheme was developed for strategies that paralleled the item classification categories. There was substantial overlap in strategies used by males and females; however, analyses of strategy types used across all items indicated that females were more likely than males to use algorithmic strategies and males were more likely than females to use insightful strategies. It should be noted that these findings constitute a generalization across a group of subjects and items, and that there were several individual instances of males using more algorithmic strategies than females and females using more insightful strategies than males.

Questionnaire data gathered from students who participated in interviews indicated a positive relationship for both males and females between SAT-mathematical performance and positive attitudes toward mathematics (e.g., liking mathematics as a subject and recognizing its usefulness to their adult lives). The use of algorithmic strategies, however, was correlated with negative attitudes toward mathematics (e.g., mathematics being difficult and not being relevant to their lives). Implications of these and other findings are discussed.

INTRODUCTION

Sex differences in mathematical performance are well documented, though the hypothesized causes of these differences are varied. To date, however, little work has been done that examines the specific *nature* of these differences—exactly what males and females do differently on measures of mathematical performance. The research presented here seeks to add to our understanding of the nature of sex differences in performance on standardized mathematics tests.

Previous research has shown that males perform better than females on the mathematical sections of many standardized achievement tests administered to adolescents. Although some studies have found that the size of average differences is small, relatively large differences (of about half a standard deviation) have been found among high achieving students (Benbow and Stanley 1980; Feingold 1988; Hyde, Fennema, and Lamon 1990). Research findings generally conclude that sex differences in mathematical performance occur among adolescent and adult populations but not among younger groups (Maccoby and Jacklin 1974; Hyde, Fennema, and Lamon 1990). These differences have been found on standardized tests but not in mathematics grades (Clark and Grandy 1984).

Recent efforts to examine sex differences in mathematical performance have attempted to validate hypotheses concerning the origins of the observed differences. Many of the studies examining the effects of background on mathematical performance have focused on the relation between sex and courses taken. It is well documented that males take more mathematics courses than females (Armstrong 1981; Benbow and Stanley 1982; Fennema and Sherman 1977). Several studies have found that controlling for course-taking substantially reduces, but does not eliminate, sex differences in performance on standardized tests (Bridgeman and Wendler 1990; Pallas and Alexander 1983).

Studies frequently conclude that differences in test performance are the result of differential performance on only certain types of items. Indeed, the bulk of the research supports a dichotomy of item types: items that are well-defined and resemble material taught in school, versus items that are ill-defined and require insightful or unusual solution strategies. Armstrong (1985) and Dossey, Mullis, Lindquist, and Chambers (1988) found that females outperform males on problems where the procedural rule is obvious, while males do better when the problem-solving strategy is less clearly defined. This dichotomy can conceivably encompass the categorization of item types showing sex differences in performance in other studies: the pure mathematics/word problem categorization (McPeek and Wild 1987; O'Neill, Wild, and McPeek 1989), the algebra/geometry categorization (Boswell 1985; Doolittle and Cleary 1987; Hudson 1986), and the algorithmic/strategic categorization (Doolittle 1987).

These results support hypotheses put forth by Kimball (1989) in an effort to explain why females tend to get better grades than males in mathematics classes but show poorer performance on mathematics tests. Kimball suggests that differences in test performance may be the result of differences in the way male and female students approach and solve problems.

Results of an earlier statistical study performed on SAT-mathematical data (Gallagher 1990) indicated that high-scoring males and females may indeed be using different solution strategies on SAT-mathematical items. In that study, a taxonomy was developed to classify mathematics items on three forms of the SAT on the basis of strategies that could be used in their solution. When Differential Item Functioning (DIF) procedures were performed on the item data for examinees scoring 650 or above, significantly more items requiring insight or estimation were flagged in favor of males than items requiring the use of standard algorithms. When self-reported course-taking in math was controlled, similar differences favoring males were found for the best-prepared group (examinees who had four years or more of high school mathematics, including some calculus).

It is hypothesized by some researchers that these sex differences in the ability to apply mathematical principles to solve unfamiliar problems are the result of males' and females' different approaches to learning mathematics (Fennema and Peterson 1985; Grieb and Easley 1984). According to this hypothesis, the learning process for males is largely independent of procedures outlined in the classroom, whereas females rely substantially more on algorithms taught in class and procedures outlined by the teacher. According to Grieb and Easley (1984), autonomous learning exhibited by males allows them to develop their own solutions to problems without relying on the procedures and algorithms taught in class. Although this hypothesis presents a reasonable explanation for the differences, there is meager empirical evidence to support it.

It appears, then, that sex differences in performance on standardized tests of mathematical ability are probably the result of a combination of background variables and differences in approaches to problem solving and learning in mathematics. Males may possess a combination of characteristics that allows them to outperform females on standardized tests of mathematical problem solving but that does not generally give them an advantage on measures such as grades in mathematics courses. It appears that some of the observed differences in high-performing students can be accounted for by different strategies used by males and females to solve particular types of problems.

To date researchers have not directly examined what males and females do on standardized tests that produces results contradictory to those provided by mathematics grades. Work examining performance on particular item types has identified content and format variables often associated with sex differences in performance, but no study has documented how these variables interact with the operations students actually use to solve mathematical problems. The present study examines sex differences in performance on SAT-mathematical items from the perspective that performance differences are the result of different strategies used by males and females to solve mathematical problems.

The purpose of this study was to further our understanding of performance differences on the SAT by fine-tuning the classification scheme for mathematics items developed by Gallagher (1990) in order to predict performance among high-ability students more accurately. This study also sought to develop a classification scheme for problem-solving strategies that parallels the item classification scheme. The strategy classification scheme was used to examine student problem solving directly, thus providing a link between quantitative analyses of response data and empirical data on how students solve problems. Finally, information on students' attitudes and academic and extracurricular activities was examined to determine whether these were related to strategy use or performance on the SAT.

The study consisted of three parts. Items were classified and analyzed in the first part of the study. In the second part, items that were identified as differentially functioning for males and females were presented to students, who solved them out loud during recorded interviews. These data were used to validate the item categories used in the analyses in Part 1 and to examine differences in the types of strategies male and female students actually used. Part 3 involved analyzing the questionnaire data obtained from the students who participated in Part 2.

PART 1: ITEM ANALYSIS

Method

The Mantel-Haenszel DIF procedure was run on subjects scoring 650 or more for five forms of the SAT. Two types of DIF statistics were used to determine differences in item performance: the Mantel-Haenszel P-DIF (MH P-DIF) and the Mantel-Haenszel D-DIF (MH D-DIF). The P-DIF statistic, based on the normally distributed percent correct (P+) distribution, is more sensitive to differences in items of middle difficulty, where small differences on the ability continuum are equivalent to a larger area under the normal curve than differences at either end of the scale. The D-DIF statistic, on the other hand, is based on delta, which is a linear scale and is, therefore, more sensitive than P-DIF to differences in items of high or low difficulty.

Examinees in the DIF analyses were drawn from one administration of each of the five forms that were analyzed: May and November 1987, May and November 1988, and May 1989. All analyses were run on subjects who scored 650 or above, which restricted the range of possible scores to 150 points as opposed to the standard 600-point range. This restriction of range in scores was reflected in relatively small DIF values. Consequently a relatively small criterion value was used to identify differentially functioning items.

Items were flagged if the MH D-DIF was greater than or equal to 0.5 and the MH P-DIF was at or above a comparable level for that scale (.05). Although higher levels of the MH D-DIF (1.0 or higher) are used operationally for flagging items on the SAT, a value of 0.5 or more is considered appropriate for research purposes. Items were classified and the average MH P-DIF and MH D-DIF statistics for item categories by sex were examined using an analysis of variance procedure.

Item Classification

Before any analyses of the DIF statistics were conducted, items in the mathematical sections of the five forms of the SAT were classified according to the types of strategies that could be efficiently used in their solution given the time constraints of the test. The classification scheme developed by Gallagher (1990) was revised and the "item content" category was eliminated, since results of previous analyses indicated that there was no relationship between item content and performance by sex for high-scoring examinees.

In order for this study to further our knowledge of problem-solving strategies used by high scorers on the SAT, it was important to have consistency in item classification across raters and across test forms. With this in mind, a reiterative process was used to develop the category descriptions listed below. A "bottom-up" approach was used to develop new categories by grouping the flagged items from the five forms. Thirty items were flagged across the five forms in the DIF analyses. Twenty-five of these items as well as two items whose MH P-DIF values were slightly below the flagging criterion (.047 and .048 as opposed to .05) were used for this purpose. These items were selected for the development of the new classification system because the analysis of item responses provided with the DIF analyses indicated a substantial sex difference (5 percent or more) in the choice of a particular distractor or in omit rates.

The first step in the process of developing new classifications was to examine the items to determine whether they could be classified according to common attributes. Items were first divided into two groups: items favoring males and items favoring females. These were then reviewed, along with an analysis of the most common wrong answer selected by each sex, to determine whether they could be grouped into subsets based on strategies that were probably used in their solution. This task was performed with the assistance of a test developer (female) who has had extensive experience in teaching mathematics, writing mathematics items, and determining examinees' solution strategies on the basis of notations written in test booklets. The following categories resulted from this review:

1. Items that do not require insight, but require fairly lengthy solutions (these items favored females).
2. Items that are too time-consuming if solved by standard methods but that can be solved quickly with insight (these items favored males).
3. Items for which a standard algorithm does not exist or that require an unusual application of a standard algorithm (these items favored males).
4. Items on which female responses indicated a misapplication of a standard algorithm (these items favored males).

Next, the same items were reviewed by a male test developer with similar background and experience to determine whether there was consensus on the groupings. This review resulted in two changes to the classification scheme: the elimination of the fourth category, since the hypothesized reason for errors was too subjective, and the inclusion of estimation as a strategy. The following three categories resulted from this review:

1. Questions that are most efficiently solved using a standard algorithmic strategy of the type generally taught in high school mathematics classes.
2. Questions for which the use of a standard algorithm would be too time-consuming and that therefore require the examinee to invent an estimation strategy.
3. Questions that must be solved using a unique algorithm invented specifically for that problem using general mathematical principles.

These three categories were then used to classify all the items on the five forms of the SAT. The two test developers who were involved in developing the categories classified the 300 items independently, and the percent agreement was calculated after all items were classified. Table 1 (first rating) shows that agreement ranged from 47 percent to 82

Table 1. Interrater Agreement on Item Classifications

Form	% Agreement	
	First Rating	Second Rating
May 1987	47	75
Nov. 1987	73	78
May 1988	62	67
Nov. 1988	73	85
May 1989	82	77

Table 2. Reliability of Final Item Classifications

Form*	% Agreement on Reclassification
May 1987	93
Nov. 1987	93
May 1988	93
Nov. 1988	95
May 1989	95

*Each form contains 60 items.

percent, with greater than 70 percent agreement on three forms.

The items on which the raters did not agree were reviewed and the classification categories were revised again to include two additional attributes: (1) specific mention of items that require the examinee to determine whether a relationship between two variables or equations is constant for all possible values of the variables, and (2) a distinction between the insightful use of a standard algorithm and a unique algorithm created specifically for a particular problem. The following four categories resulted from this final revision:

1. Questions that can be answered only by primarily algorithmic methods. The method of solution is clearly defined. These items are examples of routine textbook problems (type I).
2. Questions that can be answered using a textbook-type algorithm but that may be solved more quickly using estimation or insight (type II).
3. Questions that *require* an insightful use of an algorithm generally taught in school (type III).
4. Questions that are not generally found in textbooks and require the use of a unique algorithm developed specifically for the problem. These questions may require consideration of all cases critical to satisfying given conditions. All items whose key is "It cannot be determined by the information given" fall in this category (type IV).

Items on the five forms were reclassified using the new categories. As Table 1 (second rating) shows, the overall percent agreement between raters increased. Agreement across forms was between 67 percent and 85 percent, with 75 percent or greater agreement on four of the five forms. Items on which the raters disagreed were discussed again, and the raters reached consensus on all but six items (98 percent agreement). The six items on which the raters did not agree were not classified.

To test the reliability of the classification system, the raters were requested to reclassify the same items one month later. Agreement was between 93 and 95 percent (see Table 2). The 19 (out of 300) items for which classifications differed retained the classification assigned during the final session. To facilitate the use of the classification system in future research, examples of prototypical items from each category are provided in Appendix A.

Subjects

Subjects for this part of the study were examinees who had recently taken one of the five forms of the SAT during a regular administration. The MH DIF procedure was used to examine existing data in order to identify items on which males and females performed differently. In order to employ this procedure it was necessary to match males and females on total SAT-mathematical scores. For example, males who received a score of 600 had to be compared with females who scored 600. On average, females' scores were lower than those of males. Therefore, the lower cutoff score for subjects in this part of the analysis was based on the females' performance. Given the above requirements, the subject pool for this analysis consisted of all examinees that were high school juniors or seniors, considered English to be their first language, and scored at or above the ninety-fifth percentile for female examinees.

In addition, since previous analyses had shown that having taken calculus was related to SAT-mathematical scores (Gallagher 1990), only subjects who had taken calculus were used in the analyses. This information was obtained from the Student Descriptive Questionnaire (SDQ). The final sample, which drew subjects from all five forms, consisted of 67,887 males and 31,218 females.

The recency of the calculus courses as well as the school in which they were taken were not addressed in this analysis. Although these two factors have been shown to have some effect on SAT-mathematical performance, this information is not provided by the SDQ. Since this part of the study relied on existing data, these two factors were not controlled.

Results

Item Location and Difficulty

The new classification scheme yielded a different distribution of item types across forms than was found in previous work (Gallagher 1990). Table 3 shows the frequency distributions of item types across all five test administrations for the new classification categories. Fifty-eight percent of the items were classified as type I, and types II through IV constituted 12 to 16 percent each.

Table 3. Frequency Distributions of New Item Types across Five Test Administrations

Item Type	*Frequency*	*% of Test (rounded)*
I (algorithm)	175	58
II (algorithm or estimation)	35	12
III (insight with algorithm)	48	16
IV (logic or insight only)	36	12
Unclassified	7	4

In keeping with the findings of Gallagher (1990), the present analysis indicated that type I items made up the majority of the items at the beginning of each section. Indeed, 70 percent of the items in the first half of each section were classified as type I, whereas only 46 percent of the items in the second half of each section were type I items. Item types III and IV were more frequently found in the second half of sections, and item type II was found about equally in both halves. Since item location in the SAT is determined by difficulty, one could conclude that for the total group of examinees at any given administration, items classified as type III or IV were generally more difficult for the population of SAT takers than those classified as type I.

High scorers in the present study follow a similar pattern. Fifty percent of all type III items and 33 percent of all type IV items had percent correct (P+) values less than 0.8 (less than 80 percent of the examinees answered correctly). In contrast, only 12 percent of type I and type II items had P+ values below 0.8. Table 4 presents the mean P+ values for each type of item. Across the five forms, mean P+ values for type I and type II items are generally larger (indicating that a greater proportion of the sample answered them correctly) than those for item types III and IV. An analysis of variance was run to determine whether any of these differences were significant. The asterisks in Table 4 indicate that the mean P+ for item type I was significantly different from the mean P+ for item types III and IV.

Sex Differences

When the DIF analysis was run, a total of 30 items were flagged. All flagged items demonstrated a significant sex difference on the Mantel-Haenszel chi-square test ($p < .001$). Table 5 displays the distribution of items that were flagged favoring males or females. Nine of the 10 flagged items classified as type I (standard algorithm) favored females, and all the flagged items classified as types II, III, and IV favored males. Significantly more items were flagged favoring males than females (binomial test $p < .05$). The mean P+ values for flagged items favoring males and females were 0.68 and 0.73, respectively. This difference was not significant.

To compare performance patterns in the high-scoring group with those of the average group, the preceding analyses were also performed on the total group of students who took the SAT during the five selected administrations. Table 6 displays the distribution of items flagged for the total group of examinees using the criterion described above. More type I items were flagged for the total group than for the high-scoring group, and the total number of items favoring males and females is much more closely balanced in the total group than in the high-scoring group. Further, flagged items for the total group consisted of a different set of items than were flagged for the high-scoring group. However, the pattern of items favoring males and females within item categories in the total group is similar to that of the high-scoring group; the majority of flagged type I items favored females, and the majority of flagged type II, III, and IV items favored males. These preliminary data suggest that some of the differences found among high scorers are also present among students at lower levels.

Table 7 presents mean MH P-DIF values across all items in the analysis of the high-scoring group. A one-way analysis of variance was run on the MH P-DIF values of the

Table 4. Mean P+ Values for Item Types and Test Administrations

Item Type	*Test Administration* 5/87	11/87	5/88	11/88	5/89	*Combined*
I	.92	.90	.93	.92	.92	.92
II	.93	.85	.80	.94	.94	.89
III	.75	.70	.83	.75	.81	.78*
IV	.84	.82	.87	.85	.85	.84*

*Significantly different from type I ($p < .05$).

Table 5. Distribution of Flagged Items for High-Scoring Examinees

Item Type	*Favored Sex* *Male*	*Female*	*Total*
I	1	9	10
II	6	0	6
III	7	0	7
IV	7	0	7
Total	21	9	30

Table 6. Distribution of Flagged Items for Total Group of Examinees

Item Type	*Favored Sex* *Male*	*Female*	*Total*
I	18	27	45
II	5	1	6
III	4	2	6
IV	7	1	8
Total	34	31	65

Table 7. Mean Values of MH P-DIF for Item Types

Item Type	*Mean**	*SD*
I**	.0080	0.02
II	.0143	0.03
III	.0089	0.03
IV	.0205	0.04

*Positive values indicate that females were favored and negative values indicate that males were favored.

**Significantly different from all other types ($p < .001$)

300 items in the five forms to determine whether there was a significant effect among high scorers by item type. Since a large proportion of examinees correctly answered the majority of the mathematics items, the P+ values were generally high. Therefore, the MH P-DIF statistic was selected as the unit of analysis over the MH D-DIF statistic because it is the more conservative measure of DIF for items with high P+ values. A significant main effect was found for item type ($F = 17.01$; $p < .0001$). Tukey's post hoc test showed significant differences between type I items and all other item types.

A mean of 0 signifies that neither sex is favored, a negative mean favors males, and a positive mean favors females. Values approaching ± 0.1 are considered extreme. An examination of the means in Table 7 reveals that females performed better than males on type I items, and males performed better than females on types II, III, and IV. Sex differences in performance on type I items were significantly different from those on other item types. When these values are compared to Gallagher's (1990) mean MH P-DIF values, they are found to be more extreme and cluster more consistently by sex.

Discussion

The refinements made to the item classification scheme developed by Gallagher (1990) produced a new classification scheme that was better able to distinguish between items that favor males and those that favor females among high-scoring students. Using the refined scheme, the distribution of item types within each test section was similar to that found in the earlier study; item types III and IV were generally placed at the end of a section, indicating that they were more difficult. However, items that were flagged showed no significant difference in difficulty by item type. The distribution of items favoring males over females was also similar to that found in the earlier work; there were more than twice as many flagged items favoring males as those favoring females.

The main difference between the classification scheme used by Gallagher (1990) and the refined version used in this study is in the proportion of items falling into each category (item type) and the mean MH P-DIF values for categories. The refined scheme classified a greater proportion of items as type I than the original version did (58 percent versus 32 percent), and the MH P-DIF values clustered more consistently by sex. Mean values for item types were more extreme, and there was a significant difference between type I items that slightly favored females and type II, III, and IV items that favored males to varying degrees.

When the distribution of flagged items for the high-scoring group was compared with the distribution of items flagged for the total group, a number of similarities and differences were apparent. The two groups differed in the proportion of items that were flagged as favoring one sex over the other. In the total group, about the same number of items favored males as favored females. In contrast, many more items favored males than favored females in the high-scoring group. The two groups also differed in the distribution of flagged items across item categories. Many more type I items (algorithmic) were flagged for examinees in the total group than in the high-scoring group. Within item types, however, the ratio of items favoring males to those favoring females is similar. The majority of flagged type I items favored females, and the majority of flagged type II, III, and IV items favored males in both the total group and the high-scoring group. This suggests that problem-solving differences that appear to exist among high scorers might also be found among examinees in lower score ranges.

PART 2: PROTOCOL ANALYSIS

Score differences found for high-ability students on the mathematical sections of the SAT usually favor males, but results of the item analyses that control for total SAT-mathematical score indicate that these differences are probably due to differential performance on only certain types of items. The purpose of the second part of the study was to provide qualitative information that could clarify the findings of the statistical studies. High scorers on the SAT were interviewed to determine what strategies they were actually using on SAT-mathematical items. The hypothesis that differences in the performance of high-scoring males and females could be accounted for, in part, by differences in the solution strategies they use was examined further. Specifically, it was hypothesized that females would be more likely than males to use standard applications of school-taught algorithms, and males would be more likely than females to invent nonstandard types of solution strategies. Results of the item analyses are consistent with this hypothesis, but they lend only indirect support. The protocol analyses that follow were performed to examine students' problem-solving strategies directly in an effort to document these differences in a small sample of students.

Method

Fifty-eight subjects were asked to think aloud while solving the four types of mathematical problems identified in the item analysis. Think-aloud sessions were audiotaped. According to Ericsson and Simon (1984), think-aloud protocols or concurrent verbal reports can be considered reliable data on subjects' thought processes. The authors state:

> Our examination of two of the most vigorous challenges to the usefulness of verbal reporting leaves intact our belief that such reports—especially concurrent reports . . . of specific cognitive processes—provide powerful means for gaining information about such processes. The concurrent report reveals the sequence of information heeded by the subject without altering the cognitive process. (p. 30)

Subjects

Subjects for this part of the study were juniors and seniors from public and private high schools within half an hour's drive from Educational Testing Service (ETS) in central New Jersey. Students were recruited through mathematics teachers, department heads, or guidance counselors. The original sample of 58 was reduced to 47 due to failure of the recording device. The resulting sample consisted of 25 males and 22 females whose SAT-mathematical scores were 670 or above. This score represents approximately the ninety-fifth percentile for females in recent administrations of the test. Table 8 displays the distribution of SAT-mathematical scores for these subjects.

Instruments

The item set for the think-aloud protocols (Appendix B) consisted of 27 items flagged for sex differences in the DIF analysis. Twenty-two of these items were part of the group of items used to develop the item classification categories in the item analysis study. Items were selected on the basis of the size of the sex difference and their level of difficulty.

The items were pilot tested on subjects who were similar in age and mathematical ability to the subjects who would provide the protocols. Results of the pilot testing showed that most students were able to solve individual items in four minutes or less, and that the level of difficulty was appropriate for the group that had been selected. Items were administered in order of difficulty. Two easy items that were not scored were given as practice.

Data Collection

Students and their parents were requested to read and sign a consent form with a brief description of the research. The think-aloud protocols were collected on audiotape by a female examiner at the subject's school either during school hours or after school. An empty classroom or office was used to keep distractions at a minimum and to ensure that audiotapes were intelligible. Problems were presented one at a time, each on a full sheet of paper, and subjects were encouraged to use the extra space for scratch. Problem sheets for each student were labeled and retained to aid in transcription of the protocols. Students were told that the research they were participating in sought to examine how high-scoring students solved mathematical problems on the SAT. Subjects were instructed to read the following written instructions before the think-aloud sessions began:

Table 8. Distribution of SAT-M Scores by Sex

Score	*Males*	*Females*
670	0	2
680	2	2
690	3	1
700	1	1
710	1	4
720	2	4
730	2	1
740	2	3
750	3	2
760	2	1
770	1	0
780	3	1
790	1	0
800	2	0
Mean	740	720

> I am going to give you some math problems from the SAT. I would like you to do these problems in the same way you would if you were taking the SAT except I would like you to "think out-loud" while you solve them. Say everything that you are thinking while you are getting your answer.
>
> You will have four minutes to do each problem. Please use the extra space on each page to write any notes or equations. The first two problems will be for practice. Remember, try to do the problems in exactly the same way you would if you were really taking the SAT. After you finish a problem I will not tell you whether your answer is correct or not, but I may ask you to explain some more about how you got your answer. Do you have any questions?

No other information was given to students before or during the interview. Interviews lasted from 40 minutes to an hour. Their duration depended on the amount of time available in the student's schedule and how quickly the problems were solved. The examiner intervened only in the following circumstances: (1) If the subject paused for more than 30 seconds, the examiner asked, "What are you thinking?" (2) If the subject exceeded the time limit, the examiner informed him or her that the allotted time was up. (3) If the examiner was unable to understand how the subject solved the problem, the subject was asked, "Can you tell me how you got that?"

Strategy Classification

Strategy classification categories were developed in much the same way as the new item classification categories described earlier. Once all the think-aloud protocols were recorded and transcribed, solutions were examined within each problem. Methods of solution for each problem were listed. This resulted in a list of up to 10 strategies for each problem. Next, solution strategies for all problems were compared to determine whether there were commonalities across problems. The following eight categories of solutions were developed through this process:

1. *Algorithm:* Solutions that consist primarily of computational strategies generally taught in school. This includes computations and algebraic formulas using abstract terms or givens from the problem stem.
2. *Insight with algorithm:* Solutions that use a mathematical algorithm but are simplified or shortened due to insight, logical reasoning, or estimation. This category includes solutions for which the student realizes that it is not necessary to complete an equation or algorithm in order to choose an option.
3. *Logic, estimation, or insight:* Solutions based primarily on the application of mathematical principles or logic, either alone or in combination with estimation or insight. These solutions generally do not employ computations or algorithms, but they may include minor mental calculations. Solutions may require consideration of all cases critical to satisfying given conditions.
4. *Assigning values to variables:* Solutions achieved by assigning values to variables given in the problem stem. This includes trial-and-error solutions using random numbers and solutions for which assigned values make operations more "concrete."
5. *Plugging in options:* Solutions found by working backwards from options, systematically plugging in choices.
6. *Guessing:* Solutions based primarily on guessing. This includes choices that are based on surface characteristics of options or are unexplained (e.g., "it just looked right"). This category does not include estimations based on partial solutions.
7. *No strategy:* Examinee did not formulate a solution strategy. This category applies only to instances in which no attempt at solution was made (e.g., "I don't know how to do this"). It does *not* include items the examinee attempted but failed to solve.
8. *Misinterpretation:* This category applies only to solutions for which the examinee clearly misread or misinterpreted the item stem or diagram. It does not include faulty solutions caused by misunderstanding or misuse of mathematical terms and concepts or unfinished solutions.

Transcriptions of all solutions to each problem were coded by one rater using the eight categories listed above. If the subject changed strategy within a problem, more than one strategy was coded for that item. A second rater was then trained by the first rater to use the coding system. Training consisted of a review of the categories followed by a discussion of prototypical solutions to several problems. The second rater then practiced coding solutions on two problems. Ratings for solutions to these problems were then reviewed and discussed with the first rater. Finally, the second rater coded all solutions to 10 of the 27 problems that had not been specifically discussed with the first rater. The two problems used for practice coding were not included in the 10 problems. Coding by both raters was "blind" in that coders did not know the subject's sex. There was 82 percent agreement between the first and second raters. Disagreements were resolved through discussion to reach 100 percent agreement.

Results

Since the distribution of SAT-mathematical scores in this part of the study was somewhat higher for males than for females, two sets of analyses were run on the solution strategy data. The first set was run on the total group of 47 subjects. An additional set was run on a subsample of the total group ($n = 28$) in which males and females were matched on SAT-mathematical scores.

Strategy Use

Strategies used by students in the think-aloud protocols were categorized into one of the eight categories outlined above. Nine of the 27 items used in the think-aloud protocols were eliminated from the analyses because the protocols provided little insight into the processes that students used to solve them. Seven of these items were "quantitative comparison" items. Solutions to this item type consisted of either plugging actual values into the equations (with no indication of how the values were selected) or simply choosing an answer "because it has to be that way." Probing examinees on their solution methods did not ameliorate this situation. Clearly other methods more specifically geared to quantitative comparison items must be used to determine the strategies students use to solve these items.

The remaining pool of items consisted of 18 items: 9 type I items, 4 type II items, 3 type III items, and 2 type IV items. Since the statistical analyses (performed in the item analysis) found that only type I items (standard algorithm) were significantly different from all other types (which require either insight or the application of a novel algorithm), most analyses were performed on only two groups of items: algorithm items (type I; $n = 9$) and insight items (types II–IV; $n = 9$).

Strategy categories can also be grouped into strategies that apply standard algorithms and those that require insight. Strategy types 1 (applies a computational algorithm), 4 (substitutes values for variables), and 5 (works backwards from options) can all be grouped in an algorithm category, since they are all fairly standard computational strategies. Strategies 2 (insight with algorithm) and 3 (logic, estimation, or insight) can be grouped in an insight category, since they both require insight based on mathematical principles.

When students changed solution strategies while solving a particular item, multiple strategies were coded for that item. The incidence of multiple strategy use was fairly low, ranging from about 10 percent to 15 percent by item. How-

Table 9. Percent of Males and Females Responding Correctly by Item Type and Strategy Type for Total and Matched Groups

Total Group

	*Strategy Type**															
	1		*2*		*3*		*4*		*5*		*6*		*Algorithm***		*Insight***	
Item Type	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*
I (algorithm)	66	75	17	8	2	3	11	10	3	4	0	1	81	88	19	11
II	41	38	38	42	16	13	0	0	5	6	0	0	46	44	54	56
III	10	21	73	41	2	0	13	21	2	14	0	0	25	56	75	41
IV	0	0	21	27	49	43	26	27	0	0	5	3	26	27	70	70
Insight items (II, III, IV)													34	42	64	54

Matched Group

	*Strategy Type**															
	1		*2*		*3*		*4*		*5*		*6*		*Algorithm***		*Insight***	
Item Type	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*
I (algorithm)	71	79	13	8	3	1	11	8	3	4	0	0	84	91	16	9
II	44	41	33	46	11	5	0	0	11	8	0	0	56	49	44	51
III	4	23	83	39	0	0	8	22	4	17	0	0	17	61	83	39
IV	0	0	17	30	52	50	26	20	0	0	4	0	26	20	70	80
Insight items (II, III, IV)													33	45	63	54

*Strategies 7 and 8 were excluded because they result in incorrect answers.

**Algorithm = strategies 1, 4, and 5; insight = strategies 2 and 3; guessing = strategy 6.

ever, because it was possible that this could affect the outcome, analyses were run on two sets of data: (1) all strategies that students used in solving problems, and (2) strategies that produced an answer—or final strategies. Results of the various analyses of both sets of data showed that multiple strategies had little effect on the outcome of the analyses. Therefore, results are reported only for the data set of final strategies.

Item responses were examined using a Mantel-Haenszel chi-square to determine whether there were sex-related differences in the proportion of items answered correctly. For the total group, the only significant difference in the ratio of correct to incorrect responses was found for insight items (types II, III, and IV) [$X^2(1) = 5.7; p < .02$], where males responded correctly 80 percent of the time and females responded correctly 69 percent of the time. These results agree with the results of the statistical analyses that showed that females did better on algorithmic items (type I) and males outperformed females on insight items.

In the matched group, females outperformed males across all item types [$X^2(1) = 4.2; p < .05$]. When the items were grouped it became clear that the females' superior performance on algorithm items was the major contributor to the difference in overall performance [$X^2(1) = 9.63; p < .01$]. On type I items, females responded correctly 81 percent of the time, and males responded correctly only 62 percent of the time. There was no significant difference in performance on insight items.

Strategy categories were examined to determine whether the students' strategies were those expected to be used based on the item classifications. Table 9 displays the percent of correct responses for the total and matched groups falling in the various strategy categories. When items and strategies are grouped into algorithm and insight categories, one can see that the majority of strategies used on algorithm items (type I) were algorithmic strategies (1, 4, and 5), and the majority of strategies used on insight items (types II–IV) were insightful strategies (2 and 3).

In addition, the extent to which the model predicted strategies used on individual items was examined. Student records were analyzed to determine whether the strategies they used were those that the model predicted they would use on specific items. According to the model, item type dictates which strategies will be used to find the solution. So type I items (algorithm) would be solved by both sexes using an algorithmic strategy, but type II items (algorithm or insight) would be solved by females using an algorithmic strategy and by males using an insightful strategy. Type III items (algorithm with insight) would be solved by both sexes in the same manner, as would type IV items (insight alone).

Table 10. Males and Females Using Predicted Strategy Types on Items

Percent of Strategies Predicted	
90—	
-	
88-	M*
-	F*
86-	
-	
84-	
-	M*, M*, M*, F*
82-	F*
-	
80—	
-	
78-	
-	M**, M**, M**, M**, M**, F**, F**, F**
76-	
-	M**
74-	
-	M***, F***
72-	M**, M**, M**, M**, F**, F**, F**, F**
-	F***, F***
70—	
-	M, M, M, F, F
68-	
-	M, M, M, M, F, F
66-	
-	F
64-	M, F
-	
62-	M
-	
60—	
-	
58-	F
-	F
56-	
-	
54-	
-	M
52-	
-	
50—	F
-	

Note: Subjects in the matched group are underlined.
*$p < 0.01$
**$p < 0.05$
***$p < 0.1$

Since all subjects did not reach all items, the number of responses ranged from 10 to 18. The prediction rate ranged from a low of 50 percent (5 out of 10 items) to a high of 88 percent (14 out of 16 items), with the average rate of prediction for males and females being 73 percent and 71 percent, respectively. Table 10 displays the percent "hit rate" for the model prediction of individual male and female strategies. A binomial test was conducted for each subject to determine whether the hit rate was better than chance. The hit rate was found to be significantly greater than chance for 15 of the 25 males ($p < .01$) and 13 of the 22 females ($p < .01$).

The data also support the hypothesis that females are more likely than males to use algorithmic strategies even on items that require insight. This is most clearly seen in the grouped item and strategy categories. For the total group, females' use of algorithmic strategies was 8 percent greater than males' on algorithm items and 15 percent greater than males' on insight items. A Mantel-Haenszel chi-square test was used to determine whether these differences were significant. The distribution of algorithmic and insightful strategies was significantly different for males and females on all items [$X^2(1) = 17.5$; $p < .001$] and on insight items [$X^2(1) = 8.5$; $p < .01$]. The difference in strategy use for algorithm items was not significant for this group.

The data for the matched group also showed females using algorithmic strategies more often than males on all item types. On algorithm items females used algorithmic strategies 16 percent more often than males, and on insight items females used algorithmic strategies 14 percent more often. The chi-squares indicated that these differences were significant for all items combined [$X^2(1) = 9.5$; $p < .01$], as well as for algorithm items [$X^2(1) = 5.7$; $p < .02$] and insight items [$X^2(1) = 4.1$; $p < .05$].

Strategy use within subjects was also examined. Table 11 shows the distribution of students by the percentage of problems solved using algorithmic strategies. Sixty percent of the females used algorithmic strategies at least 70 percent of the time, but fewer than one-quarter of the males used algorithmic strategies to this extent. Half of the males used algorithmic strategies on 50 percent or less of the problems, but only about one-tenth of the females fell into this category. Table 11 also indicates that some students tend to use primarily one type of strategy or the other; others seem to use both strategies fairly equally.

Discussion

An analysis of strategies used by high-scoring students on SAT-mathematical items supported the item categories developed in the item analyses. Correct responses to items that were classified type I (requiring a standard algorithmic solution) were generated with standard solution strategies taught in school more than 80 percent of the time. Correct responses on items classified type II, III, or IV (all containing an insight component) tended to be produced with more unusual, insightful solution strategies.

Further, these results indicate that there is a sex difference in the way these high-scoring students solved particular types of problems. In the analyses of the total group, significant differences were found in strategies used on both algorithm and insight items, where females were more likely to use algorithmic strategies than males. In keeping with the results of the item analysis, however, there was virtually no difference in the proportion of correct answers given on items requiring the application of standard strategies taught in school. The difference in performance appears on items that require some type of insight or estimation, where males outperformed females.

Table 11. Distribution of Subjects by Strategy Use

% of Problems Solved with Algorithm	*Total Group*		*Matched Group*	
	Males	*Females*	*Males*	*Females*
83	0	1	0	1
78	1	1	1	1
72*	3	5	3	4
67	1	3	0	2
61	2	3	1	3
56	2	2	1	1
50*	9	3	4	1
44	3	2	1	1
39	1	1	0	0
33	1	1	1	0
28	2	0	2	0

*If examinees complete all items, the model predicts that females would use 72 percent algorithmic strategies and males would use 50 percent algorithmic strategies.

In the matched group, performance differences were in the same direction, except that females outperformed males. As in the total group, females tended to use algorithmic strategies more often than males on all item types. When performance differences were examined by item type, it became clear that the female advantage was gained on algorithm items, where there was a significant difference in performance, rather than on insight items. One can see that the patterns for males and females in both the matched group and the total group are similar. Relative to the males, females did better on algorithm items and worse on insight items. Conversely, relative to females, males did better on insight items and worse on algorithm items.

PART 3: ANALYSIS OF QUESTIONNAIRE DATA

Students who participated in the protocol analysis study were given a questionnaire at the end of each think-aloud session. The purpose of collecting the questionnaire data was to determine whether there was a relationship among high-scoring examinees' performance on the SAT, the types of strategies they used, their attitudes toward mathematics, their test-taking strategies, and their academic and extracurricular activities.

Method

The questionnaire (Appendix C) was composed of three sections: (1) a section regarding attitudes toward mathematics, based on Fennema and Sherman's (1976) Mathematics Attitude Scale; (2) a list of items describing various test-taking strategies; and (3) questions taken from the Student Descriptive Questionnaire (SDQ) for the SAT. Questionnaires were given to students at the end of think-aloud interviews. Each student was asked to complete the questionnaire during free time and return it in a self-addressed postage-paid envelope that was provided.

Statistical analyses were run to determine the relationship between questionnaire items and students' SAT-mathematical scores. In addition, analyses were run to determine whether questionnaire items were related to the students' problem-solving strategies used during think-aloud sessions. This was done by correlating questionnaire responses with the percent of strategies each student used that were algorithmic (i.e., strategy 1 solutions).

Results

The return rate for questionnaires was high. Seventy-six percent of the students interviewed completed and returned the questionnaire. Of the 58 students that were interviewed, 21 females and 23 males returned completed questionnaires. Appendix D provides details of students' background and individual questionnaire items correlating with SAT-mathematical performance and use of algorithmic strategies.

The majority of the questionnaire respondents were white (70 percent) 17-year-old twelfth graders (73 percent). Most of them had taken the SAT within the last year (91 percent), and most had taken it twice (59 percent). These students generally prepared for the SAT on their own, studying at home rather than taking test preparation courses at school or elsewhere (68 percent). Most students rated themselves as being in the highest 10 percent in mathematical ability (86 percent) and at least above average in science (84 percent) and writing ability (77 percent). Students generally reported that both of their parents held a bachelor's or graduate degree (95 percent of fathers and 82 percent of mothers) and that their family income was generally greater than $70,000 (60 percent). First choices for college major were spread over 14 different fields, with the most popular choices for females being economics, banking, or business (n = 5) and English or creative writing (n = 3). The most popular choice of major for males was engineering (n = 5). Finally, half of the students responded that they planned to complete degrees at the doctoral level.

When Pearson product-moment correlations were run, different items showed significant correlations with SAT-mathematical performance and algorithmic strategy use (see Appendix D for specific items and correlation coefficients). The magnitude of correlations was moderate, ranging from .37 to .76. For the total group of respondents, correlations indicated a positive relationship between performance on the SAT and self-confidence in mathematics and a liking for the subject. Other factors that correlated positively with SAT-mathematical performance were persistence in solving problems and the strategy of leaving several problems unanswered. Factors that correlated negatively with SAT-mathematical performance for the total group were guessing, the notion that mathematical problems can always be solved quickly, and the number of high school English courses taken.

When correlations were examined separately by sex, different items were significantly correlated with SAT-mathematical performance for males than for females. For males, persistence in solving mathematical problems, a sense that mathematics will be useful in adult life, and the strategy of leaving several items unanswered all correlated positively with SAT performance. Factors that were negatively correlated with SAT performance for males were guessing, years of foreign language study in high school, and years of participation in government or political activities. Positive correlations were found for females between SAT-mathematical performance and liking mathematics as a subject and grades in English classes. Females' SAT-mathematical scores correlated negatively with the number of high school English and natural sciences courses taken.

A different set of items correlated with the use of algorithmic strategies. For the total group, the use of algorithms correlated positively with students not wanting other students to know they did well in mathematics classes, ratings of mathematics as a difficult subject, and belief that mathematics is not relevant to the student's life. In addition, the number of courses taken in high school mathematics correlated positively with the use of algorithms in the total group. Items that correlated negatively with the use of algorithms were statements about the usefulness of mathematics, liking others to know they do well in mathematics, and the strategy of working a problem out before looking at the options.

When correlations were run by sex, results for males showed positive correlations between the use of algorithms and the number of high school English courses and the number of years studying algebra. Algorithmic strategy use in males also correlated with the notions that they did not like others to know they were good at mathematics and that girls who liked mathematics were a bit peculiar. The notion that mathematics is a useful subject to study correlated negatively with males' use of algorithms.

For females, positive correlations were found between the use of algorithms and the notion that mathematics is difficult and irrelevant to their lives, as well as the number of high school foreign language courses taken. Negative correlations for females were found between the use of algorithms and being thought of as smart in mathematics, working a problem out first and then selecting from the options, and having taken a test preparation course outside of school.

Discussion

Although selection of the sample of students to be interviewed in the protocol analyses was not specifically geared toward obtaining a sample that was representative of all students who scored above 650 on the mathematical sections of the SAT, statistics for ethnicity, income, and parents' education for the group of students who were selected come fairly close to those of the total group of college-bound seniors who scored above 650 on the mathematical sections of the SAT in 1989 (Robertson 1990). High-scoring college-bound seniors were 80 percent white as opposed to the present sample, which was 70 percent white. Seventy percent of the high-scoring population in 1989 reported that their fathers had received at least a bachelor's degree, as opposed to 95 percent of the current sample. Finally, a notable proportion of high scorers from 1989 reported that their parents' income was above $70,000 (30 percent) as opposed to 60 percent in the present group. The difference in this last factor can most likely be attributed to the limited geographic area from which the current sample was selected.

Results of the questionnaire analyses indicate that, even among this high-scoring group, the students who do best on the SAT are those who like mathematics as a subject, are confident in their mathematical ability, and feel that mathematics is a useful subject that is relevant to their present and future lives. These students also take fewer courses in verbal areas such as English and foreign languages.

Fewer items correlated with SAT-mathematical performance when analyses were broken down by sex. It is likely that this is the result of the relatively small sample size. As one would expect, for both males and females, attitudes that related positively to SAT-mathematical performance were attitudes that were positive toward mathematics: persistence in problem solving and the relevance of mathematics to adult life correlated positively for males, as did liking mathematics for females. Items that correlated negatively were taking courses in verbal areas (English and foreign languages) or in natural sciences.

It is interesting to note that many of the items that demonstrated a positive relationship to the use of algorithmic strategies were items that indicated a negative attitude toward mathematics: that mathematics is difficult and not useful, and not wanting to be thought of as "smart" in mathematics. This suggests that students who have less affinity for mathematics are the ones who rely most heavily on algorithmic strategies; others who like the subject better and have more confidence are more likely to use more creative strategies.

It is also interesting that algorithmic strategy use correlated positively with both the number of courses taken in verbal areas such as English and foreign languages and the number of courses taken in mathematics, and that SAT-mathematical performance did not correlate positively with either.

CONCLUSIONS

This study examined performance on mathematics items requiring different types of solution strategies in order to determine whether high-scoring males and females differed in their use of standard algorithms and insight. The results of these analyses help to clarify the nature of sex-related differences in performance on the SAT. Findings described here offer direct support for the notion that at least a portion of the differences among high scorers can be attributed to differences in strategy use. Females in this group appear to depend more heavily than males on standard algorithmic strategies that are generally taught in the classroom; males are more apt to use insight in their solutions. In addition, the students (both male and female) who used more algorithmic strategies tended to rate mathematics as more difficult and less relevant to their lives.

The following discussion is restricted to high-ability students only. Hyde, Fennema, and Lamon's (1990) work indicates that performance differences are substantially larger in high-ability groups, suggesting that different factors could be involved at various levels of the ability continuum. Although preliminary data from the comparison of item types flagged for high scorers and the total group suggest that gender-related differences in mathematics performance at upper score levels and for the average group may have some factors in common, the present study focused primarily on high-scoring students. For this reason, results of this study may not be generalized to other levels. In addition, since items in the quantitative comparison format were omitted from the analyses, results of this study do not apply to this type of item. The results of this study are summarized below.

Summary of Results

The item classification scheme developed by Gallagher (1990) was refined, resulting in a more accurate prediction of sex differences in performance on mathematical sections of the SAT. When DIF procedures were performed on item data for examinees scoring above 650, the majority of items that were flagged as favoring males required the use of mathematical insight, whereas all the items flagged as favoring females required standard algorithmic solutions. This supports the findings of previous work and the notion that males do better on items requiring insightful solutions and females tend to do better on items requiring standard algorithmic solutions.

An analysis of item data using the revised item classification scheme indicated that less than half of the items on the SAT require insight, but that these items are generally more difficult than items requiring the application of a standard algorithm. It is interesting to note, however, that even though the items that favored females were generally easier than those favoring males, there was no significant difference in the difficulty of items flagged as differentially favoring males and females.

A classification scheme was developed for strategies that was parallel to the item classification categories. Solution strategies were obtained from think-aloud protocols. Strategies were analyzed to determine whether students actually used the types of strategies hypothesized in the item classification scheme. In general, the strategies used were of the type predicted. An analysis of strategy types supported the notion that females are more likely than males to use standard algorithmic strategies, and males are more likely than females to use insightful or novel solution strategies.

Analyses of questionnaire data eliciting information about students' attitudes, test-taking strategies, and academic and extracurricular activities indicated a positive relationship for both males and females between SAT-mathematical performance and positive attitudes toward mathematics (e.g., a liking for mathematics as a subject and recognizing that it will be useful in their adult lives). The use of algorithmic strategies, on the other hand, was positively correlated with rather negative attitudes toward mathematics (e.g., mathematics is difficult and not relevant to their adult lives). Few relationships were found between SAT-mathematical performance or strategies and academic and extracurricular activities. The only significant correlations were the negative relationships of English and foreign language courses to SAT-mathematical performance and the positive relationship between the use of algorithmic strategies and courses taken in mathematics, English, and foreign languages.

Discussion

The null results for item difficulty on flagged items in Part 1 of the study combined with the sex differences found for strategies in Part 2 seem to support the differential strategy notion. If males and females used essentially equivalent strategies but males were just better, then the difference in performance would be related to item difficulty. The fact that differences in male and female performance were not related to item difficulty suggests that they are using different processes to solve the problems.

Since females were more likely than males to use algorithmic solution strategies, which require the systematic application of rules or formulas, they tended to do somewhat better than males on items that required this type of solution. An example of this is illustrated by problem 7 in Appendix B, on which females outperformed males. On this particular problem, most students counted units around the perimeter to determine which rectangle was larger. Females tended to count out the perimeter of *all* rectangles, whereas males tended to count only a few. Males who got this item wrong generally did so because they did not systematically count the perimeters of all rectangles; they counted only the perimeters of the rectangles that appeared to be the largest (e.g., B and C) and selected among them. So, on this particular item, those who were more thorough and consistently applied the counting algorithm rather than looking for shortcuts had an advantage.

Problem 12 in Appendix B is an example of a problem for which use of insight provides an advantage. There are two ways to solve this problem: using a computational strategy or using a logical strategy. The computational strategy is fairly complex and requires some time and effort to set up equations; indeed, many who chose this strategy failed to set up equations that worked. Using logic, however, one can see that more than half of the mixture must be the cheaper coffee (espresso), since the price per pound for the mixture is less than half the sum of the price per pound of each type. Using this information, there is only one possible option that could be correct, since all the others are greater than 25 (or half of the 50 pounds of the mixture). On this type of problem, students who tended to look for shortcuts using their knowledge of mathematical principles had an advantage; they saved the time it would take to set up and solve the equations and avoided the pitfalls of extensive computation.

The sex-related differences in solution strategies that were found here provide insight into the discrepancy between mathematics test scores and class grades. If classroom-based tests require the application of solution strategies that have been taught (familiar, well-defined strategies), then students who are more apt to spontaneously use these strategies (females) will outperform students who are less likely to use them. On the other hand, standardized tests that contain items whose solutions have not been explicitly taught would yield differences favoring males, who are more apt to use less conventional solution strategies.

The differences in solution strategies found here also suggest that differences in self-confidence that have been found in previous work could hold true for this group of subjects as well. The fact that females spent more time solving problems and were more likely to use school-taught algorithms over more insightful methods reflects a more conservative approach that could be born of a lack of self-confidence. Indeed, correlations between strategy use and questionnaire data on attitudes indicate that for the group of subjects interviewed, those who used algorithms most frequently found mathematics to be a difficult subject that had little relevance to their lives. The approach that many males take, on the other hand, relies more heavily on insight or estimation. Confidence in one's own ability could lead a student to use these types of shortcut strategies that ultimately require less time and effort.

Several hypotheses speculate on the origins of these sex differences. Both the autonomous learning behavior hypothesis and the novelty/familiarity hypotheses reviewed by Kimball (1989) provide reasonable explanations for the differences in strategy use found here. Although there is little or no empirical evidence for these hypotheses, they suggest that mathematics teachers might be able to influence how students go about solving problems.

According to these two hypotheses, females are more likely than males to use solution strategies provided by the teacher and, as a result, are less likely to do well on novel item types for which they have not learned a specific solution strategy. If these theories are correct, then teachers can help reduce differences in the ways students solve problems by discussing several different ways of solving problems and by introducing different types of problems.

According to *Professional Standards for Teaching Mathematics*, recently published by the National Council of Teachers of Mathematics (NCTM 1991), the goal of teaching is to help students develop "mathematical power." The NCTM defines mathematical power as:

> the ability to explore, conjecture, and reason logically; to solve nonroutine problems; to communicate about and through mathematics; and to connect ideas within mathematics and between mathematic and other intellectual activity. Mathematical power also involves the development of personal self-confidence and a disposition to seek, evaluate, and use quantitative and spatial information in solving problems and in making decisions. Students' flexibility, perseverance, interest, curiosity and inventiveness also affect the realization of mathematical power. (p. 1)

All these characteristics would contribute to increased performance on items requiring unusual solution strategies and to an increased understanding of the relevance of mathematics to other areas of one's life.

The NCTM also notes that current teaching practices often do not foster these types of abilities in students. It suggests that several shifts in the environment of the mathematics classroom need to take place in order for all students to be able to develop mathematical power. Four of the five recommended shifts appear to relate directly to the differences in problem solving found in the current study:

1. Toward logic and mathematical evidence as verification—away from the teacher as the sole authority for right answers.
2. Toward mathematical reasoning—away from merely memorizing procedures.
3. Toward conjecturing, inventing, and problem solving—away from an emphasis on mechanistic answer finding.
4. Toward connecting mathematics, its ideas, and its applications—away from treating mathematics as a body of isolated concepts and procedures.

If the changes in the teaching of mathematics that are recommended in the NCTM standards can be implemented in most classrooms, it is possible that the types of sex differences in problem solving that were found here will be greatly reduced.

This report has not examined the cause of the differences reported here. Hypotheses that speculate on the origin of sex differences in mathematics performance have been presented, and suggestions for changes in teaching strategies have been suggested. Differences in the use of problem-solving strategies in mathematics may be the result of a long sequence of events that affect the sexes differently. High-scoring females, as a group, seem to be somewhat more conservative in their strategies, sticking to methods they were taught in school. This may be caused by a lack of confidence or interest, or because of the way they learned and think about mathematics. Factors such as these could conceivably be affected by changes in teaching strategies such as those recommended by the NCTM. Few studies to date have directly examined the impact of social learning on sex differences among high-ability students or how great an effect socialization has on these differences in mathematics. These questions and questions of how these differences can be eradicated need to be examined in future research.

Suggestions for Future Research

Many researchers have speculated on the origins of sex differences in mathematics, but these theories lack empirical foundation. Future research should attempt to provide a link between speculations such as the autonomous learni hypothesis and empirical data. Before information concerning the nature of these differences in problem solving can be useful to educators, it will be necessary to link strategy differences to causal factors that might influence change. Research needs to be done that will link strategy differences to other factors such as self-confidence or early learning that have been documented in the literature.

Although the results of this study suggest that females are more likely to use standard algorithmic strategies and males are more likely to use insightful strategies, conclusions can be drawn only for high-ability groups. Further research is needed to determine whether this phenomenon exists at other ability levels and whether there are other, equally important differences that are more apparent at lower levels.

One final note of caution in interpreting the findings of this study: Although sex differences in solution strategies were found, there was substantial overlap in the strategies used by both sexes. Results of any study examining cognitive differences between the sexes should be understood in the context of this overlap—as generalizations about fairly small differences. In most cases, the overlap between the sexes is greater than the differences between them.

REFERENCES

Armstrong, J. M. 1981. "Achievement and Participation of Women in Mathematics. Results of Two National Surveys." *Journal of Research in Mathematics Education* 12(5): 356–72.

Armstrong, J. M. 1985. "A National Assessment of Participation and Achievement of Women in Mathematics." In *Women and Mathematics: Balancing the Equation*, edited by S. F. Chipman, L. R. Brush, and D. M. Wilson (pp. 59–94). Hillsdale, N.J.: Erlbaum.

Benbow, C. P., and J. C. Stanley. 1980. "Sex Differences in Mathematical Ability: Fact or Artifact?" *Science* 210:1262–64.

Benbow, C. P., and J. C. Stanley. 1982. "Consequences in High School and College of Sex Differences in Mathematical Reasoning Ability: A Longitudinal Perspective." *American Educational Research Journal* 19(4):598–622.

Benbow, C. P., and J. C. Stanley. 1983. "Sex Differences in Mathematical Reasoning Ability: More Facts." *Science* 222:1029–31.

Block, J. H. 1984. *Sex Role Identity and Ego Development*. San Francisco: Jossey-Bass.

Boswell, S. L. 1985. "The Influence of Sex-Role Stereotyping on Women's Attitudes and Achievement in Mathematics." In *Women and Mathematics: Balancing the Equation*, edited by S. F. Chipman, L. R. Brush, and D. M. Wilson (pp. 175–98). Hillsdale, N.J.: Erlbaum.

Bridgeman, B., and C. Wendler. 1990. *Prediction of Grades in College Mathematical Courses as a Component of SAT-M Placement Validity*. College Board Report No. 89-9, New York: College Entrance Examination Board.

Brophy, J. 1985. "Interactions of Male and Female Students with Male and Female Teachers." In *Gender Influences in Classroom Interaction*, edited by L. C. Wilkinson and C. B. Marrett (pp. 115–42). Orlando, Fla.: Academic Press.

Brush, L. R. 1985. "Cognitive and Affective Determinants of Course Preferences and Plans." In *Women and Mathematics: Balancing the Equation*, edited by S. F. Chipman, L. R. Brush, and D. M. Wilson (pp. 123–50). Hillsdale, N.J.: Erlbaum.

Chodorow, N. 1974. "Family Structure and Feminine Personality." In *Women, Culture, and Society*, edited by M. Z. Rosaldo and L. Lamphere. Stanford: Stanford University Press.

Clark, M. J., and J. Grandy. 1984. *Sex Differences in the Academic Performance of Scholastic Aptitude Test Takers*. College Board Report No. 84-8, New York: College Entrance Examination Board.

Connor, J. M., and L. A. Serbin. 1985. "Visual-Spatial Skill: Is It Important for Mathematics? Can It Be Taught?" In *Women and Mathematics: Balancing the Equation*, edited by S. F. Chipman, L. R. Brush, and D. M. Wilson (pp. 151–74). Hillsdale, N.J.: Erlbaum.

Doolittle, A. E. 1987. *Gender Differences in Performance on Mathematics Achievement Items*. ACT Research Report 87-16.

Doolittle, A. E., and T. A. Cleary. 1987. "Gender-Based Differential Item Performance in Mathematics Achievement Items." *Journal of Educational Measurement* 24:157–66.

Dossey, J. A., I. V. S. Mullis, M. M. Lindquist, and D. L. Chambers. 1988. *The Mathematics Report Card: Are We Measuring Up? Trends and Achievement Based on the 1986 National Assessment*. Princeton, N.J.: The Nation's Report Card, NAEP, Educational Testing Service.

Dweck, C. S. 1986. "Motivational Processes Affecting Learning." *American Psychologist* 41:1040–48.

Eccles, J. (Parsons), T. Adler, and J. L. Meece. 1984. "Sex Differences in Achievement: A Test of Alternate Theories." *Journal of Personality and Social Psychology* 46:26–43.

Eccles, J. S., and P. Blumenfeld. 1985. "Classroom Experiences and Student Gender: Are There Differences and Do They Matter?" In *Gender Influences in Classroom Interaction*, edited by L. C. Wilkinson and C. B. Marrett (pp. 79–114). New York: Academic Press.

Elliot, E. S., and C. S. Dweck. 1988. "Goals: An Approach to Motivation and Achievement." *Journal of Personality and Social Psychology* 54:5–12.

Ericsson, K. A., and H. A. Simon. 1984. *Protocol Analysis: Verbal Reports as Data*. Cambridge, Mass.: MIT Press.

Ethington, C. A., and L. M. Wolfle. 1984. "Sex Differences in a Causal Model of Mathematics Achievement." *Journal for Research in Mathematics Education* 15:361–77.

Feingold, A. 1988. "Cognitive Gender Differences Are Disappearing." *American Psychologist* 43(2):95–103.

Fennema, E. 1974. "Mathematics Learning and the Sexes: A Review." *Journal for Research in Mathematics Education* 5:126–29.

Fennema, E., and J. Sherman. 1976. "Fennema-Sherman Mathematics Attitude Scales: Instruments Designed to Measure Attitudes toward the Learning of Mathematics by Females and Males." *JSAS Catalog of Selected Documents in Psychology* 6:31 (ms. no. 1225).

Fennema, E., and J. Sherman. 1977. "Sexual Stereotyping and Mathematics Learning." *The Arithmetic Teacher* 24(5): 369–72.

Fennema, E., and P. Peterson. 1985. "Autonomous Learning Behavior: A Possible Explanation of Gender-Related Differences in Mathematics." In *Gender Influences in Classroom Interaction*, edited by L. C. Wilkinson and C. B. Marrett (pp. 17–35). New York: Academic Press.

Fennema, E., and L. A. Tartre. 1985. "The Use of Spatial Visualization in Mathematics by Girls and Boys." *Journal for Research in Mathematics Education* 16(3):184–206.

Gallagher, A. M. 1990. *Sex Differences in the Performance of High-Scoring Examinees on the SAT-M*. College Board Report No. 90-3. New York: College Entrance Examination Board.

Grieb, A., and J. Easley. 1984. "A Primary School Impediment to Mathematical Equity: Case Studies in Rule-Dependent Socialization." In *Advances in Motivation and Achievement: Women in Science*, edited by M. W. Steinkamp and M. L. Maehr (pp. 317–62). Greenwich, Conn.: JAI Press.

Hilton, T. L., and G. W. Bergland. 1974. "Sex Differences in Mathematics Achievement: A Longitudinal Study." *Journal of Educational Research* 67:231–37.

Hudson, L. 1986. "Item-Level Analysis of Sex Differences in Mathematics Achievement Test Performance." *Dissertation Abstracts International* 47(2) (order no. DA8607283).

Hyde, J. S. 1981. "How Large Are Cognitive Gender Differences?" *American Psychologist* 36(8):892–901.

Hyde, J. S., E. Fennema, and S. J. Lamon. 1990. "Gender Differences in Mathematical Performance: A Meta-Analysis." *Psychological Bulletin* 107(2):139–55.

Johnson, E. S. 1984. "Sex Differences in Problem Solving." *Journal of Educational Psychology* 76(6):1359–71.

Kimball, M. M. 1989. "A New Perspective on Women's Math Achievement." *Psychological Bulletin* 105:198–214.

Licht, B. G., and C. S. Dweck. 1983. "Sex Differences in Achievement Orientations: Consequences for Academic Choices and Attainments." In *Sex Differentiation and Schooling*, edited by M. Marland (pp. 72–97). London: Heinemann.

Linn, M. C., and A. C. Petersen. 1986. "A Meta-Analysis of Gender Differences in Spatial Ability: Implications for Mathematics and Science Achievement." In *The Psychology of Gender: Advances through Meta-Analysis*, edited by J. S. Hyde and M. C. Linn (pp. 67–101). Baltimore: Johns Hopkins University Press.

Maccoby, E. E., and C. N. Jacklin. 1974. *The Psychology of Sex Differences*. Stanford: Stanford University Press.

Marshall, S. P. 1984. "Sex Differences in Children's Mathematics Achievement: Solving Computations and Story Problems." *Journal of Educational Psychology* 76(2):194–204.

McPeek, W. M., and C. L. Wild. 1987. *Characteristics of Quantitative Items that Function Differently for Men and Women*. Paper presented at the annual meeting of the American Psychological Association, New York.

National Council of Teachers of Mathematics. 1991. *Professional Standards for Teaching Mathematics*. Reston, Va.: NCTM.

O'Neill, K., C. L. Wild, and W. M. McPeek. 1989. *Gender-Related Differential Item Performance on Graduate Admissions Tests*. Paper presented at the annual meeting of the American Psychological Association, San Francisco.

Pallas, A. M., and K. L. Alexander 1983. "Sex Differences in Quantitative SAT Performance: New Evidence on the Differential Coursework Hypothesis." *American Educational Research Journal* 20(2):165–82.

Robertson, N. 1990. *1989 College-Bound Seniors High-Scoring on SAT-V (> 600) and on SAT-M (> 650), with Contrast to Total Group of 1989 College-Bound Seniors*. Unpublished manuscript.

Senk, S., and Z. Usiskin. 1983. "Geometry Proof Writing: A New View of Sex Differences in Mathematics Ability." *American Journal of Education* 91:187–201.

Sherman, J. 1980. "Mathematics, Spatial Visualization, and Related Factors: Changes in Girls and Boys, Grades 8–11." *Journal of Educational Psychology* 72(4):476–82.

APPENDIX A: PROTOTYPICAL ITEMS FOR ITEM CLASSIFICATION CATEGORIES

Type I

If n oranges cost p dollars, at this rate how many dollars will 5 oranges cost, in terms of n and p?

(A) $5np$

(B) $\frac{n}{5p}$

(C) $\frac{p}{5n}$

(D) $\frac{5n}{p}$

(E) $\frac{5p}{n}$

If $m^2 - n^2 = 4k^3$, $m - n = 12k$, and $k \neq 0$, what is $m + n$ in terms of k?

(A) $\frac{3}{k^2}$

(B) $\frac{k^2}{3}$

(C) $3k^2$

(D) $4k^3 + 12k$

(E) $48k^4$

Type II

Each plant of a certain variety yields 50 seeds in the early fall and then dies. Only 40 percent of these seeds produce plants the following summer and the remainder never produce plants. At this rate, a single plant yielding seeds in 1986 will produce how many plants as descendants in 1989?

(A) 60

(B) 400

(C) 8,000

(D) 16,000

(E) 32,000

A blend of coffee is made by mixing Colombian coffee at \$8 a pound with espresso coffee at \$3 a pound. If the blend is worth \$5 a pound, how many pounds of the Colombian coffee are needed to make 50 pounds of the blend?

(A) 20

(B) 25

(C) 30

(D) 35

(E) 40

Type III

Ms. Smith spent 20 percent of the money she had in her purse on lunch and 75 percent of what was left after lunch on groceries. If she then had \$5.00 left, how much had she spent on lunch?

(A) \$10.00

(B) \$8.50

(C) \$6.00

(D) \$5.00

(E) \$4.50

SUMMARY DIRECTIONS FOR COMPARISON QUESTIONS
Answer: A if the quantity in Column A is greater; B if the quantity in Column B is greater; C if the two quantities are equal; D if the relationship cannot be determined from the information given.

Column A		Column B
	$a + b + c = 15$ $b + c + e = 21$ $a + e = 12$	
The average (arithmetic mean) of a, b, c, and e		7

Type IV

If x, $\frac{1}{x}$, y, $\frac{1}{y}$, and $\frac{1}{z}$ are integers, which of the following could NOT be a value of $x + y + z$?

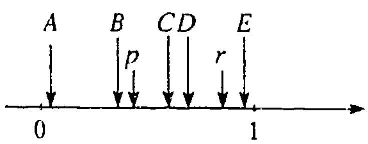
(A) 4

(B) 3

(C) 1

(D) -1

(E) -3



On the number line above, which of the lettered arrows could be pointing to the product $p \times r$?

(A) A

(B) B

(C) C

(D) D

(E) E

APPENDIX B: PROBLEMS USED FOR THINK-ALOUD PROTOCOLS

SUMMARY DIRECTIONS FOR COMPARISON QUESTIONS
Answer: A if the quantity in Column A is greater; B if the quantity in Column B is greater; C if the two quantities are equal; D if the relationship cannot be determined from the information given.

1. (item type IV)

Column A | Column B

$6z < 7y$

z | y

2. (item type I)

If $m^2 - n^2 = 4k^3$, $m - n = 12k$, and $k \neq 0$, what is $m + n$ in terms of k?

(A) $\frac{3}{k^2}$

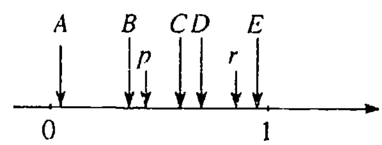
(B) $\frac{k^2}{3}$

(C) $3k^2$

(D) $4k^3 + 12k$

(E) $48k^4$

3. (item type IV)



On the number line above, which of the lettered arrows could be pointing to the product $p \times r$?

(A) A
(B) B
(C) C
(D) D
(E) E

4. (item type I)

Mary's present age is half Paul's age. In 4 years, Paul will be n years old. In terms of n, how old is Mary now?

(A) $\frac{n}{2} - 2$

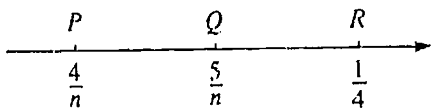
(B) $\frac{n}{2}$

(C) $\frac{n}{2} + 2$

(D) $\frac{n}{2} + 4$

(E) $2n + 4$

5. (item type I)



If $PQ = QR$ on the number line above, what is the length of PR?

(A) $\frac{1}{12}$

(B) $\frac{1}{9}$

(C) $\frac{1}{8}$

(D) $\frac{1}{6}$

(E) $\frac{3}{16}$

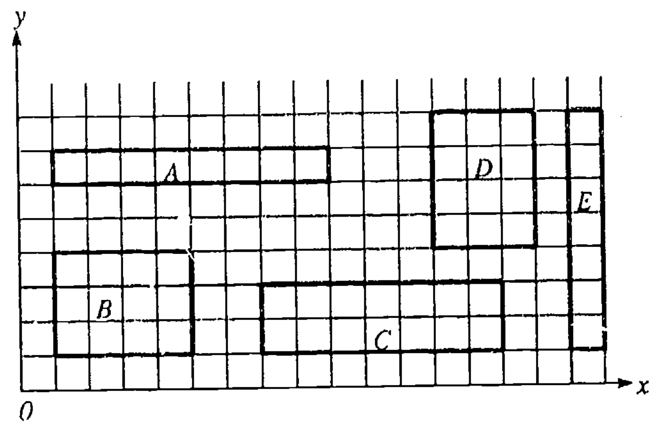
SUMMARY DIRECTIONS FOR COMPARISON QUESTIONS

Answer: A if the quantity in Column A is greater;
B if the quantity in Column B is greater;
C if the two quantities are equal;
D if the relationship cannot be determined from the information given.

6. (item type IV)

Column A		Column B
	$x > y > z > 0$	
$\frac{x}{y}$		$\frac{y}{z}$

7. (item type I)



In the scale drawings above, each unit along the x-axis represents 2 meters and each unit on the y-axis represents 1 meter. Which of the rectangles above represents the rectangle with the greatest perimeter?

(A) A
(B) B
(C) C
(D) D
(E) E

8. (item type IV)

If x, $\frac{1}{x}$, y, $\frac{1}{y}$, z, and $\frac{1}{z}$ are integers, which of the following could NOT be a value of $x + y + z$?

(A) 4

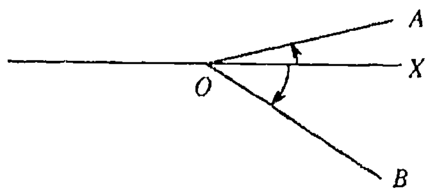
(B) 3

(C) 1

(D) -1

(E) -3

9. (item type I)



Segments OA and OB, shown in the figure above, start on OX at the same time, and revolve simultaneously in the plane in opposite directions about point O. If OA revolves at 2° per second and OB revolves at 6° per second, in how many seconds after they start will OA and OB first meet?

(A) $22\frac{1}{2}$

(B) 45

(C) 50

(D) 90

(E) It cannot be determined from the information given.

10. (item type I)

In $\triangle ABC$, D is the midpoint of AC, and another point, E, is on AC such that $BE \perp AC$. Which of the following must be true?

(A) $\angle ABD = \angle DBC$

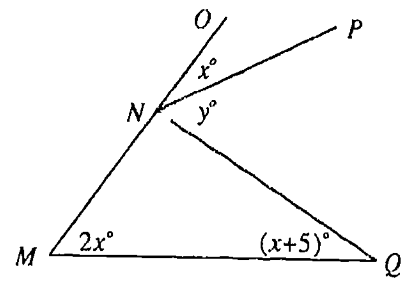
(B) $\angle EBD = \frac{1}{4} \angle ABC$

(C) $\angle BAD = \angle BDA$

(D) The length of $AC >$ the length of BC

(E) The length of $BD >$ the length of BE

11. (item type I)



In the figure above, N lies on line segment MO. Which of the following gives y in terms of x?

(A) $2x$

(B) $2x + 5$

(C) $3x + 5$

(D) $90 - x$

(E) $180 - 3x$

12. (item type II)

A blend of coffee is made by mixing Colombian coffee at \$8 a pound with espresso coffee at \$3 a pound. If the blend is worth \$5 a pound, how many pounds of the Colombian coffee are needed to make 50 pounds of the blend?

(A) 20

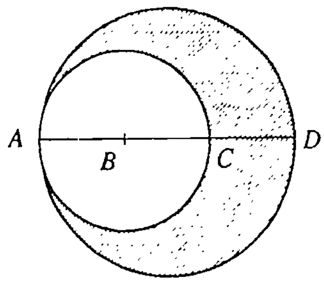
(B) 25

(C) 30

(D) 35

(E) 40

13. **(item type I)**



In the figure above, AC and AD are diameters of the small and large circles, respectively. If $AB = BC = CD$, what is the ratio of the area of the shaded region to the area of the smaller circle?

(A) 1 : 1

(B) 3 : 2

(C) 4 : 3

(D) 5 : 4

(E) 9 : 4

14. **(item type IV)**

In plane P, lines l and m are parallel to each other and are 6 inches apart. How many lines in plane P are parallel to l and m and are twice as far from one of the two given lines as from the other?

(A) None

(B) One

(C) Two

(D) Four

(E) More than four

SUMMARY DIRECTIONS FOR COMPARISON QUESTIONS

Answer: A if the quantity in Column A is greater;
B if the quantity in Column B is greater;
C if the two quantities are equal;
D if the relationship cannot be determined from the information given.

15. **(item type IV)**

Column A		Column B
	$x - 1 > 0$ and $x - 1 < 0$	
x		0

16. **(item type IV)**

Column A		Column B
	$x > 0$	
The average (arithmetic mean) of x, $2x$, and 30		The average (arithmetic mean) of x and $2x$

17. **(item type III)**

An airplane traveled a distance d in t hours, where $t > 1$, and arrived one hour late. The airplane would have arrived on time if it had traveled at what rate per hour?

(A) $t - 1$

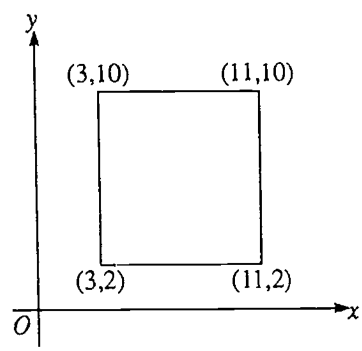
(B) $\frac{d}{t} - 1$

(C) $\frac{d}{t} + 1$

(D) $\frac{d}{t - 1}$

(E) $\frac{d}{t + 1}$

18. **(item type II)**



In the figure above, a line segment joining the point (3, 10) and which of the following points on the square will separate the square into two regions whose areas are in the ratio of 7 to 1?

(A) (11, 3)

(B) (11, 4)

(C) (11, 6)

(D) (11, 7)

(E) (11, 8)

SUMMARY DIRECTIONS FOR COMPARISON QUESTIONS

Answer: A if the quantity in Column A is greater;
B if the quantity in Column B is greater;
C if the two quantities are equal;
D if the relationship cannot be determined from the information given.

19. (item type III)

Column A		Column B
	$a + b + c = 15$ $b + c + e = 21$ $a + e = 12$	
The average (arithmetic mean) of a, b, c, and e		7

20. (item type I)

If the average (arithmetic mean) of six numbers is -6, and the sum of four of the numbers is 20, what is the average of the other two numbers?

(A) 7

(B) 8

(C) -8

(D) -28

(E) -32

21. (item type III)

Ms. Smith spent 20 percent of the money she had in her purse on lunch and 75 percent of what was left after lunch on groceries. If she then had \$5.00 left, how much had she spent on lunch?

(A) \$10.00

(B) \$8.50

(C) \$6.00

(D) \$5.00

(E) \$4.50

22. (item type II)

Each plant of a certain variety yields 50 seeds in the early fall and then dies. Only 40 percent of these seeds produce plants the following summer and the remainder never produce plants. At this rate, a single plant yielding seeds in 1986 will produce how many plants as descendants in 1989?

(A) 60

(B) 400

(C) 8,000

(D) 16,000

(E) 32,000

23. (item type I)

If n oranges cost p dollars, at this rate how many dollars will 5 oranges cost, in terms of n and p?

(A) $5np$

(B) $\frac{n}{5p}$

(C) $\frac{p}{5n}$

(D) $\frac{5n}{p}$

(E) $\frac{5p}{n}$

24. (item type II)

x increased by 10% of x yields y.
y decreased by 50% of y yields z.
z increased by 40% of z yields w.

According to the statements above, w is what percent of x?

(A) 10%

(B) 33%

(C) 77%

(D) 81%

(E) 100%

SUMMARY DIRECTIONS FOR COMPARISON QUESTIONS

Answer: A if the quantity in Column A is greater;
B if the quantity in Column B is greater;
C if the two quantities are equal;
D if the relationship cannot be determined from the information given.

25. (item type IV)

Column A | Column B

Line segments RS and TU are parallel and have the same length. M is the midpoint of RS.

Length of MT | Length of MU

26. (item type III)

The average (arithmetic mean) of 5 integers is greater than 27. If the average of the first 4 integers is 22, what is the *least* possible value of the 5th integer?

(A) 32

(B) 33

(C) 47

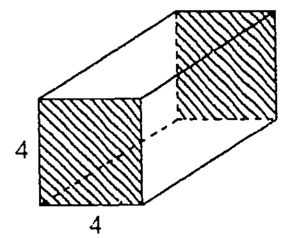
(D) 48

(E) 49

SUMMARY DIRECTIONS FOR COMPARISON QUESTIONS

Answer: A if the quantity in Column A is greater;
B if the quantity in Column B is greater;
C if the two quantities are equal;
D if the relationship cannot be determined from the information given.

27. (item type III)



Column A | Column B

Points P, Q, R, and S (not shown) are the centers of the four unshaded rectangular faces of the rectangular solid.

The area of square $PQRS$ | 16

APPENDIX C: ATTITUDE AND BACKGROUND QUESTIONNAIRE

On the following pages is a series of statements. There are no correct answers for these statements. They have been set up in a way which permits you to indicate the extent to which you agree or disagree with the ideas expressed. Suppose the statement is:

I like mathematics.

As you read the statement, you will know whether you agree or disagree. Indicate the extent to which you agree or disagree with each statement by circling the appropriate number on the 5 point scale. The scale values are 1 strongly agree, 2 somewhat agree, 3 not certain, 4 somewhat disagree, and 5 strongly disagree.

Do not spend much time with any statement, but be sure to answer every statement.

There are no "right" or "wrong" answers. The only correct responses are those that are true for you. Whenever possible, let the things that have happened to you help you make a choice.

Test-taking Strategies for SAT-Math

PLEASE CIRCLE A NUMBER WHICH BEST DESCRIBES YOUR RESPONSE.

	strongly agree				strongly disagree
1. Before I begin a section of the SAT I calculate the maximum amount of time I can spend on each problem. If I don't get an answer in that amount of time I skip the problem.	1	2	3	4	5

2. I don't think about how much time I spend on each problem, I just try to go as quickly as I can.	1 2 3 4 5
3. I try to answer every problem as I come to it.	1 2 3 4 5
4. First I answer all the problems that look easy, then I go back and do the harder ones.	1 2 3 4 5
5. I usually have plenty of time to check any answers I am unsure about.	1 2 3 4 5
6. When I come to problems that look time consuming, I can usually find another way to answer them that takes less time.	1 2 3 4 5
7. I always try to solve problems using formulas I learned in school.	1 2 3 4 5
8. I usually answer all of the problems on the test.	1 2 3 4 5
9. I usually leave several problems unanswered.	1 2 3 4 5
10. I usually guess at the answer when I can eliminate one of the options.	1 2 3 4 5
11. I only guess at the answer when I can eliminate three of the options.	1 2 3 4 5
12. I don't ever guess.	1 2 3 4 5
13. I usually look at the options immediately after reading the problem.	1 2 3 4 5
14. I usually try to work the problem out and then find an answer that matches mine.	1 2 3 4 5
15. I find it easier to solve SAT problems when I write down every step.	1 2 3 4 5
16. I find it easier to solve most SAT problems in my head.	1 2 3 4 5
17. I always check my work even if it looks right.	1 2 3 4 5
18. I only check my work when something looks wrong.	1 2 3 4 5

Attitudes

PLEASE CIRCLE A NUMBER WHICH BEST DESCRIBES YOUR RESPONSE.

1. Generally I have felt secure about attempting mathematics.	1 2 3 4 5
2. I would expect a woman mathematician to be a masculine type of person.	1 2 3 4 5
3. I'll need mathematics for my future work.	1 2 3 4 5
4. I don't like people to think I'm smart in math.	1 2 3 4 5
5. I like math puzzles.	1 2 3 4 5
6. I am sure I could do advanced work in mathematics.	1 2 3 4 5
7. Mathematics is for men; arithmetic is for women.	1 2 3 4 5
8. Studying mathematics is just as appropriate for women as for men.	1 2 3 4 5
9. I study mathematics because I know how useful it is.	1 2 3 4 5
10. It wouldn't bother me at all to take more math courses.	1 2 3 4 5
11. Mathematics is enjoyable and stimulating to me.	1 2 3 4 5
12. In terms of my adult life it is not important for me to do well in mathematics in high school.	1 2 3 4 5
13. It's hard to believe a female could be a genius in mathematics.	1 2 3 4 5
14. Girls who enjoy studying math are a bit peculiar.	1 2 3 4 5
15. I'd be happy to get top grades in mathematics.	1 2 3 4 5

16. I would trust a woman just as much as I would trust a man to figure out important calculations.	1	2	3	4	5
17. Knowing mathematics will help me earn a living.	1	2	3	4	5
18. I haven't usually worried about being able to solve math problems.	1	2	3	4	5
19. When a math problem arises that I can't immediately solve, I stick with it until I have the solution.	1	2	3	4	5
20. Mathematics is of no relevance to my life.	1	2	3	4	5
21. I think I could handle advanced mathematics.	1	2	3	4	5
22. It would be really great to win a prize in mathematics.	1	2	3	4	5
23. I see mathematics as a subject I will rarely use in my daily life as an adult.	1	2	3	4	5
24. Mathematics is a worthwhile and necessary subject.	1	2	3	4	5
25. I almost never get shook up during a math test.	1	2	3	4	5
26. If I had good grades in math, I would try to hide it.	1	2	3	4	5
27. Once I start trying to work on a math puzzle, I find it hard to stop.	1	2	3	4	5
28. I can get good grades in mathematics.	1	2	3	4	5
29. Even though I study, math seems unusually hard for me.	1	2	3	4	5
30. Males are not naturally better than females in mathematics.	1	2	3	4	5
31. I'm not the type to do well in math.	1	2	3	4	5
32. I'll need a firm mastery of mathematics for my future work.	1	2	3	4	5
33. I am usually at ease during math tests.	1	2	3	4	5
34. When a question is left unanswered in math class, I continue to think about it afterward.	1	2	3	4	5
35. Formal mathematics has little or nothing to do with everyday problem solving.	1	2	3	4	5
36. I have a lot of self-confidence when it comes to math.	1	2	3	4	5
37. I get a sinking feeling when I think of trying hard math problems.	1	2	3	4	5
38. Being regarded as smart in mathematics would be a great thing.	1	2	3	4	5
39. Math problems are always solved in less than 10 minutes if they are solved at all.	1	2	3	4	5
40. I will use mathematics in many ways as an adult.	1	2	3	4	5
41. I am usually at ease in math classes.	1	2	3	4	5
42. Only geniuses are capable of discovering or creating mathematics.	1	2	3	4	5
43. People would think I was some kind of a nerd if I got A's in math.	1	2	3	4	5
44. I am challenged by math problems that I can't understand immediately.	1	2	3	4	5
45. It is feminine for a girl to ask for help on a math problem.	1	2	3	4	5

Student Background

1. Indicate the total number of years of **high school** courses (in grades 9 through 12) you have taken in each of the subjects listed below. If you have not taken any course in a subject mark "0". If one or more of the courses is Advanced Placement, accelerated or honors, circle "H" next to the number of years.

Arts and Music (for example, art, music, art history, dance, theater)	____	H
English (for example, composition, grammar, or literature)	____	H
Foreign and Classical Languages	____	H
Mathematics	____	H
Natural Sciences (for example, biology, chemistry, or physics)	____	H
Social Sciences and History (for example, history, government, or geography)	____	H

In questions 2–5, using the same guidelines as in question 1, indicate the **total** number of years you have taken the specific courses listed.

2. **Foreign and Classical Languages**

French ____

German ____

Greek ____

Hebrew ____

Italian ____

Latin ____

Russian ____

Spanish ____

Other language courses ____

3. **Mathematics**

Algebra ____

Geometry ____

Trigonometry ____

Precalculus ____

Calculus ____

Computer Math ____

Other mathematics courses ____

4. **Natural Sciences**

Biology ____

Chemistry ____

Geology or related Earth or Space Sciences ____

Physics ____

Other science courses ____

5. **Social Sciences and History**

U.S. History ____

U.S. Government or Civics ____

European History ____

World History or Cultures ____

Ancient History .... ___

Anthropology .... ___

Economics .... ___

Geography .... ___

Psychology .... ___

Sociology .... ___

Other social sciences or history courses .... ___

6. Please enter the **average grade** for all courses you have already taken in each subject.

Arts and Music .... ___

English .... ___

Foreign and Classical Languages .... ___

Mathematics .... ___

Natural Sciences .... ___

Social Sciences and History .... ___

For questions 7 through 9, please check any of the following areas which were covered in your high school courses and related activities out of class. Check the areas you have studied or participated in (you may mark more than one in each subject area).

7. **English** course work or experience

American Literature .... ___

British Literature .... ___

Composition .... ___

Grammar .... ___

Literature of a country other than the U.S. or Britain .... ___

Literature of different historical periods .... ___

Speaking and listening skills .... ___

English as a second language .... ___

8. **Art and Music** course work or experience

I have had no course work or experience in this area .... ___

Acting or the production of a play .... ___

Art history or art appreciation .... ___

Dance .... ___

Drama or theater for appreciation .... ___

Music history, theory, or appreciation .... ___

Photography or filmmaking .... ___

Studio art and design .... ___

9. **Computer** course work or experience

I have had no course work or experience in this area .... ___

Computer literacy, awareness or appreciation .... ___

Data processing .... ____

Computer programming (BASIC, COBAL, FORTRAN, PASCAL, etc.) .... ____

Use of the computer to solve math problems .... ____

Use of the computer to solve problems in the natural sciences .... ____

Use of the computer in English courses .... ____

Word processing (use of the computer in writing letters or preparing papers) .... ____

10. **Please indicate your cumulative grade point average** for all academic subjects in **high school** (circle one).

A+	(97–100)	C+	(77–79)
A	(93–96)	C	(73–76)
A−	(90–92)	C−	(70–72)
B+	(87–89)	D+	(67–69)
B	(83–86)	D	(65–66)
B−	(80–82)	E or F	(below 65)

11. **List the grades in which you participated in the activities listed below.** If you held a **major office or position of leadership** in an activity (for example, class president, varsity team captain, officer of a statewide organization), **circle "C"** next to the grade number. Remember to include activities and accomplishments that are not school sponsored as well as your extracurricular activities.

If you have received an award or **special recognition for achievement** in an activity (for example, school prize for music or writing, varsity letter, regional science fair prize, state orchestra), **circle "A"** next to the grade number.

Academic honor society .... ________ C A

Art activity .... ________ C A

Athletics: Intramural, junior varsity, or community .... ________ C A

Athletics: Varsity or amateur-level .... ________ C A

Career-oriented activity (for example, Future Teachers of America, Future Farmers of America) .... ________ C A

Community or service activity (for example, volunteer work, neighborhood clean-up or patrol group, Scouting, Key Club) .... ________ C A

Computer activity (for example, a user's group, computer club, learning to use a computer on your own) .... ________ C A

Dance activity .... ________ C A

Debating or public speaking .... ________ C A

Ethnic or cross-cultural activity (for example, Black student organization, Hispanic club, international folk dancing) .... ________ C A

Foreign exchange or study abroad program .... ________ C A

Foreign language activity .... ________ C A

Government or political activity (for example, student government, honors council, working on a political campaign, human rights or civil rights activity in your community) .... ________ C A

Journalism or literary activity (for example, creative writing, yearbook, school newspaper, community newspaper) .... ________ C A

Junior Reserve Officers Training .... ________ C A

Music: Instrumental (for example, high school band, community orchestra, solo work) .... ______C A

Music: Vocal (for example, glee club, chorus, solo work) .... ______C A

Religious activity or organization .... ______C A

Science or mathematics activity (for example, math club, ecology or environmental group, science fair project) .... ______C A

School-spirit activity (for example, cheerleading, drill team) .... ______C A

Theater activity (for example, community or school production, acting, stage crew) .... ______C A

Work: Cooperative work program .... ______C A

Work: Part-time job, not school related .... ______C A

Other activity not listed (list below) .... ______C A

____C A

I have not participated in any of the above activities .... ______C A

12. **What is the highest level of education you plan to complete beyond high school?** (Mark only one.)

Specialized training or certificate program .... ____

Two-year associate of arts or sciences degree .... ____

Bachelor's degree (such as BA or BS) .... ____

Master's degree (such as MA, MBA, or MS) .... ____

Doctoral or related degree (such as PhD, JD, MD, DVM) .... ____

Other .... ____

Undecided .... ____

13. What is your first choice of college major? (list below)

____

14. How certain are you about your first choice of major?

Very certain .... ____

Fairly certain .... ____

Not certain .... ____

15. List up to two other majors or areas of study that interest you.

1) ____

2) ____

16. Mark each subject area in which you plan to apply for advanced placement, credit by examination, or exemption from courses.

Art .... ____

Biology .... ____

Chemistry .... ____

Computer Science .... ____

English .... ____

Foreign Languages .... ____

Humanities .... ____

Mathematics .... ____

Music .... ____

Physics .... ____

Social Studies .... ____

I don't plan to apply for exemption from these courses .... ____

17. Mark each activity you may want to take part in while in college.

Art .... ____

Athletics: Intramural sports .... ____

Athletics: Varsity sports .... ____

Community or service organization .... ____

Cooperative work or internship program .... ____

Dance .... ____

Debating or public speaking .... ____

Departmental organization (club within my major) .... ____

Drama or theater .... ____

Environmental or ecology activity .... ____

Ethnic activity .... ____

Foreign study or study abroad program .... ____

Fraternity, sorority, or social club .... ____

Honors program or independent study .... ____

Journalism or literary activity .... ____

Music: Instrumental performance .... ____

Music: Vocal performance .... ____

Religious activity .... ____

Reserve Officers Training Corps (ROTC, AFROTC, or NROTC) .... ____

Student government .... ____

18. How do you compare with other people your own age in the following three areas of ability?

Mathematical ability

____ Among the highest 10 percent ____ Above average

____ Average ____ Below average

Scientific ability

____ Among the highest 10 percent ____ Above average

____ Average ____ Below average

Writing ability

____ Among the highest 10 percent ____ Above average

____ Average ____ Below average

19. How would you describe yourself? (Mark all that apply.)

_____ American Indian or Alaskan native

_____ Asian, Asian American, or Pacific Islander

_____ Black or African American

Hispanic background:

_____ Mexican American or Chicano

_____ Puerto Rican

_____ Latin American, South American, Central American, or other Hispanic

_____ White

_____ Other

20. Indicate the highest level of education completed by your father (or male guardian) and your mother (or female guardian). Circle "F" to indicate father and "M" to indicate mother.

_____ Grade school F M

_____ Some high school F M

_____ High school diploma or equivalent F M

_____ Business or trade school F M

_____ Some college F M

_____ Associate or two-year degree F M

_____ Bachelor's or four-year degree F M

_____ Some graduate or professional school F M

_____ Graduate or professional degree F M

21. What was the approximate combined income of your parents before taxes last year?

_____ Less than $10,000

_____ About $10,000 to $20,000

_____ About $20,000 to $30,000

_____ About $30,000 to $40,000

_____ About $40,000 to $50,000

_____ About $50,000 to $60,000

_____ About $60,000 to $70,000

_____ More than $70,000

22. Please list your present grade and age.
Grade _________ Age _________

23. When was the last time you took the SAT?

Year _________ Spring Fall (circle one).

24. How many times have you taken the SAT? _________

Please write "yes" or "no" after each of the following questions.

25. Did you take a test preparation course outside of school before taking the SAT (for example, the Princeton Review, or Kaplan)?

26. Did you take a test preparation course provided by your school?

27. Did you practice for the test on your own?

APPENDIX D: BACKGROUND VARIABLES AND CORRELATIONS BETWEEN SAT-M OR STRATEGY AND QUESTIONNAIRE ITEMS

Background Variables for Subjects Participating in Protocol Study

ETHNICITY

Race	Males	Females
Asian	5	5
White	16	16
Other	2	0
Total	23	21

AGE

Age	Males	Females
15	0	1
16	4	1
17	13	17
18	6	2
Total	23	21

GRADE

Grade	Males	Females
10	0	1
11	5	6
12	18	14
Total	23	21

TEST PREPARATION

Type	Males	Females
Outside School	7	6
At School	3	3
At Home	15	15
Total*	25	24

* Categories are not mutually exclusive

SELF-RATINGS OF ABILITY

Mathematics ability	Males	Females
Top 10%	20	18
Above Avg.	3	3
Total	23	21

Science ability	Males	Females
Top 10%	13	10
Above Avg.	7	7
Average	3	4
Total	23	21

PARENTS' EDUCATION

Father	Males	Females
Business/Trade School	1	0
Some College	1	0
B.A. (4-yr. degree)	8	4
Some Grad. School	3	1
Grad./Prof. Degree	10	16
Total	23	21

Mother	Males	Females
H.S. Diploma	0	1
Business/Trade School	1	1
Some College	3	1
B.A. (4-yr. degree)	8	9
Some Grad. School	2	2
Grad./Prof. Degree	8	7
Total	22	21

INCOME

	Males	Females
$10,000-$20,000	1	0
$20,000-$30,000	1	0
$30,000-$40,000	1	0
$40,000-$50,000	1	0
$50,000-$60,000	2	0
$60,000-$70,000	0	1
> $70,000	10	16
Total	16	17

HIGHEST LEVEL OF EDUCATION PLANNED

Level	Males	Females
B.A./B.S.	0	2
M.A./M.B.A./M.S.	9	5
Ph.D./J.D./M.D.	12	10
Undecided	2	5
Total	23	22

COLLEGE MAJOR

Major	Males	Females
Architecture	0	1
Biochemistry	2	0
Bio./Earth Sciences	1	1
Chemistry	1	1
Economics/Business	0	5
Engineering	5	2
English/Writing	1	3
History	0	1
International Rel.	1	1
Languages	0	1
Mathematics	1	1
Physics	2	0
Pre-medicine	2	0
Total	16	17

Pearson Correlation Coefficients of Questionnaire Items with SAT-M Score

	Total Group	*Males*	*Females*
Test-Taking Strategies*			
I usually leave several problems unanswered.	+.53 $p < .01$	+.63 $p < .02$	–
I usually guess at the answer when I can eliminate one of the options.	–.46 $p < .01$	–.70 $p < .01$	–
Attitudes			
Generally I have felt secure about attempting mathematics.	+.39 $p < .04$	–	–
It wouldn't bother me at all to take more math courses.	+.45 $p < .02$	–	–
Mathematics is enjoyable and stimulating to me.	+.45 $p < .02$	–	+.53 $p < .05$
When a math problem arises that I can't immediately solve, I stick with it until I have the solution.	+.48 $p < .01$	+.61 $p < .02$	–
I have a lot of self-confidence when it comes to math.	+.47 $p < .01$	–	–
Math problems are always solved in less than 10 minutes if they are solved at all.	–.42 $p < .03$	–	–
I will use mathematics in many ways as an adult.	–	+.60 $p < .03$	–
Background Variables*			
Number of high school courses in English.	–.37 $p < .05$	–	–.54 $p < .04$
Number of high school courses in foreign and classical languages.	–	–.57 $p < .04$	–
Number of high school courses in natural sciences.	–	–	–.73 $p < .01$
Grades in English.	–	–	+.52 $p < .05$
Years of participation in government or political activities.	–	–.79 $p < .02$	–

* High values indicate strong agreement with the statement.

40

Pearson Correlation Coefficients of Questionnaire Items with Algorithmic Strategy Use

	Total Group	Males	Females
Test-Taking Strategies*			
I usually try to work the problem out and then find an answer that matches mine.	$-.37$ $p < .05$	–	$-.51$ $p < .05$
Attitudes*			
I don't like people to think I'm smart in math.	$+.39$ $p < .04$	–	–
I study mathematics because I know how useful it is.	$-.50$ $p < .01$	$-.54$ $p < .05$	–
Girls who enjoy studying math are a bit peculiar.	–	$+.59$ $p < .03$	–
Knowing mathematics will help me earn a living.	$-.39$ $p < .04$	–	–
Mathematics is of no relevance to my life.	$+.50$ $p < .01$	–	$+ 52$ $p < .05$
Mathematics is a worthwhile and necessary subject.	–	$-.57$ $p < .04$	–
If I had good grades in math I would try to hide it.	$+.55$ $p < .01$	$+.58$ $p < .04$	–
Even though I study, math seems unusually hard for me.	$+.43$ $p < .03$	–	$+.53$ $p < .05$
Being regarded as smart in mathematics would be a great thing.	$-.51$ $p < .01$	–	$-.76$ $p < .01$
Background Variables*			
Number of high school courses in mathematics.	$+.52$ $p < .01$	–	–
Years of high school algebra.	–	$+.63$ $p < .03$	–
Number of high school courses in English.	–	$+.62$ $p < .02$	–
Number of high school courses in foreign and classical languages.	–	–	$+.59$ $p < .03$
Having taken a test preparation course outside of school.	–	–	$-.57$ $p < .03$

* High values indicate strong agreement with the statement.